Insect Molecular Biology (2005) **14**(5), 547–553

Inactivated *mariner*-like elements (MLE) in pink bollworm, *Pectinophora gossypiella*

J. Wang*, R. Staten†, T. A. Miller‡ and Y. Park*

**Department of Entomology, Waters Hall, Kansas State University, Manhattan, KS, USA;* †*USDA-APHIS-PPQ-CPHST, Decision Support and Pest Management Systems Laboratory, Phoenix, AZ, USA; and* ‡*Department of Entomology, University of California, Riverside, CA, USA*

Abstract

We isolated multiple copies of *mariner*-like element (MLE) from the pink bollworm (PBW), *Pectinophora gossypiella*, a key lepidopteran cotton pest. Although all the MLE sequences contain multiple mutations accumulated in their transposase coding region, the consensus sequence revealed a putative ancestral transposase encoding 339 amino acid residues with a D,D(34)D motif, and 36 bp inverted terminal repeats, belonging to the *cecropia* subfamily, and most similar to the MLE found in *Antheraea* species. Examining six different pink bollworm populations, we conclude that the MLE in PBW described in this study are ancient and are undergoing the process of accumulating inactivating mutations. This conclusion is supported by the patterns of polymorphisms revealed by genomic Southern hybridization, transposable element displays, and sequences from multiple MLE.

Keywords: *mariner*, transposable element, TE display, vectorette PCR, inverse PCR.

Introduction

Transposable elements (TE) are found in the genomes of nearly all eukaryotes and make up a significant portion of the eukaryotic genome. About half of the human genome is composed of TE-derived sequences (Lander *et al*., 2001). Recent genome-wide sequence analyses of fruitfly, *Drosophila melanogaster*, and malaria mosquito, *Anopheles gambiae*, found more than 22% and 16% of the genome sequence, respectively, was composed of TE-derived sequences (Holt *et al*., 2002; Kapitonov & Jurka, 2003). The transposable element functions as an important component in evolution as a unique mutagen for beneficial, detrimental and neutral mutations of the host.

Mariner, a class II transposable element, was first discovered in *Drosophila mauritiana* as the cause of somatic genetic instability of the *peach* allele of the *white* locus (Jacobson *et al*., 1986). DNA sequences homologous to the *mariner*, *mariner-like elements* (MLE), were discovered in the moth *Hyalophora cecropia* (Lidholm *et al*., 1991) and, subsequently, in a wide range of organisms from fungi, plants, invertebrates and vertebrates (Robertson, 1993; Langin *et al*., 1995; Robertson & Zumpano, 1997; Jarvik & Lark, 1998; Mandrioli, 2000). The MLE are characterized by about 1300 bp in length, with ~30 bp inverted terminal repeats (ITR) and a single open reading frame (ORF) encoding a transposase of about 350 amino acids. Among the full-length MLE isolated from various organisms to date, however, the majority are non-functional due to accumulation of mutations through a process referred to as 'vertical inactivation' (Lohe *et al*., 1995).

The pink bollworm (PBW), *Pectinophora gossypiella* (Saunders) [Lepidoptera: Gelechiidae], is a key cotton pest. Germ line transformation of this insect, mediated by a *piggyBac* transposable element, has been successful (Peloquin *et al*., 2000). Transgenic PBW are being considered as a means to improve the Sterile Insect Technique (SIT) program that is currently used to control this insect. A PBW strain carrying a stable insertion for a visible diagnostic marker, green fluorescent protein, may be used in the SIT (Peloquin *et al*., 2000). An autocidal biological control strategy was demonstrated in *D. melanogaster* by using a strain carrying a conditional lethal mutant *Notch* gene (Fryxell & Miller, 1995). Thomas *et al*. (2000) also demonstrated a population collapse of *D. melanogaster* after the conditional expression of a lethal gene, in a strategy they called Release of Insects Carrying a Dominant Lethal (RIDL). Although biotechnology using TE to generate novel tools in pest management is developing rapidly, knowledge on the current status and evolutionary history of TE in pest populations is limited. We surveyed different populations of PBW for the presence of MLE. We report finding multiple copies of MLE in these different populations of PBW.

doi: 10.1111/j.1365-2583.2005.00585.x

Received 3 February 2005; accepted after revision 4 May 2005. Correspondence: Yoonseong Park, Department of Entomology, 123 Waters Hall, Kansas State University, Manhattan, KS66506, USA. Tel.: +1 785 532 5763; fax: +1 785 532 6232; e-mail: ypark@ksu.edu

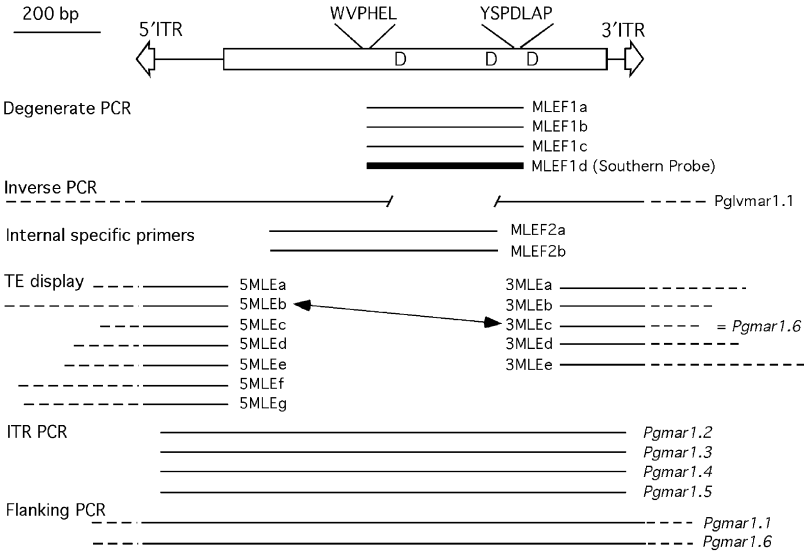

Figure 1. Diagram showing the locations of MLE clones used in the study. The top diagram shows overall structure of a typical *mariner* element. Arrows indicate inverted terminal repeats at the ends. The transposase open reading frame is boxed with conserved amino acid sequences for the degenerate primers used for initial MLE screening. Conserved three aspartic acids (D) for the D,D(34)D signature are shown in the box. Solid lines indicate clones obtained in this study. Methods used for isolating the DNA fragments are on the left of diagram. Dotted lines are the sequences for the flanking regions. The bold line is the clone used for Southern hybridization. *Pglvmar1.1* is interrupted in the middle because the clone was obtained by inverse PCR. The arrow connecting two TE display fragments is for contiguity of the clones identified in the flanking sequence PCR.

Results and discussion

Multiple copies of mariner-like elements in pink bollworm

We identified multiple copies of MLE in PBW, belonging to the subfamily *cecropia* (Fig. 1). The surveys were based on degenerate PCR using the primers MAR124F and MAR276R that were originally designed on the highly conserved amino acid sequences of the transposase (Robertson, 1993); thus, they were expected to amplify equally all MLE families. These primers have been used repeatedly for surveys of MLE in other species. Surveys relying upon degenerate PCR may be biased, however, depending on various degrees of divergence among the sequences at the primer annealing sites in distinct types of MLE. Our efforts to avoid this bias included varying PCR conditions for low stringency, and cloning and sequencing multiple clones. In this study, the degenerate PCR, presented by a total 18 clones from five different individuals (Fig. 1), identified four distinct sequences (MLEF1a to MLEF1d) belonging only to the *cecropia* subfamily, implying that the majority of MLE copies in PBW are from the *cecropia* subfamily. Multiple copies of full-length MLE were identified by a sequential approach: degenerate PCR, inverse PCR, TE display, ITR PCR, and Flanking PCR (Fig. 1 and Experimental procedures). An ancestral amino acid sequence for the PBW MLE transposase (Fig. 2A) was reconstructed based on the consensus of the five different MLE sequences (Supplementary material Fig. S1). The putative ancestral transposase encodes 339 amino acid residues with a D,D(34)D structural motif, and 36 bp inverted terminal repeats (Fig. 2B). Phylogenetic analysis supported grouping the PBW MLE in the *cecropia* subfamily with the bootstrapping value 100 in 1000 replications (Fig. 3).

The phylogram supports grouping the MLE in PBW (Family Gelechioidea) with those in *Antheraea* species (Family Saturniidae) (Fig. 3A), whereas the MLE in *Hyalophora* and *Attacus* (Family Saturniidae) and *Bombyx*, which are taxonomically closer to the *Antheraea*, are in a separate clade within the *cecropia* MLE subfamily (Fig. 3). The evolutionary patterns within the *cecropia* subfamily indicate that vertical inheritance of the MLE from a common ancestor of *Antheraea* and PBW is unlikely. As an alternative, the pattern supports horizontal transfer of the MLE, explaining the phenomenon in which the distribution and evolutionary pattern of the transposable elements is not congruent with the evolutionary relationship among the host species. The possible horizontal transfer of MLE is not surprising. Transphyla horizontal transfer has been suggested between *Attacus atlas* and *Fungia* sp. *kusabiraishi* (Fig. 2B), which just differ from each other by two nucleotide substitutions without amino acid alteration (Nakajima *et al*., 2002).

Vertical inactivation of MLE

All of the PBW MLE sequences for the region encoding the transposase revealed disrupted ORFs by the presence of multiple stop codons, deletions and insertions resulting in frame shifts. Vertical inactivation (Lohe *et al*., 1995) is a mechanism by which mutations accumulate on the active TE, and abolish or reduce the mobilization events. Patterns of polymorphism within the PBW MLE in the limited range of survey in this study suggest an explosive increase in the copy number at an early stage of the invasion, followed by vertical inactivation. Random distribution of nucleotide sequence polymorphisms in the coding region is presented by non-significant differences in the numbers of polymorphic sites in the 1st, 2nd and 3rd bases of the amino acid codons (132, 118, 140 polymorphic sites, respectively, see Supplementary material, Figs S1 and S2) with the random pattern of distribution in the deletions (Fig. S2). The random accumulation of mutations suggests

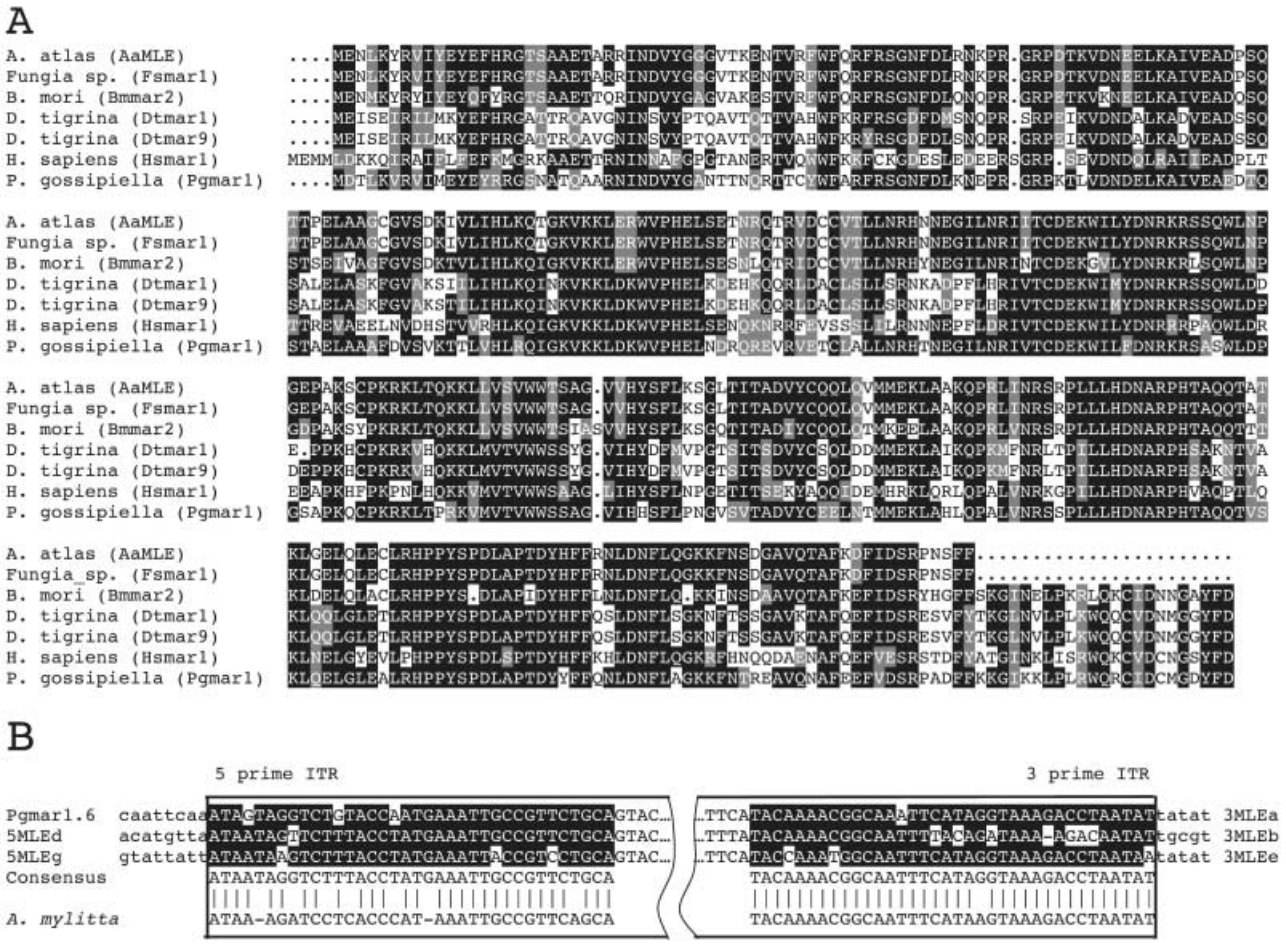

Figure 2. Boxshade of transposase amino acid sequence alignment for *cecropia* subfamily of MLE (A) and nucleotuide sequence in inverted terminal repeats (B). Black and grey boxes are for identical and similar amino acids (or identical nucleotide sequence) in 50% majority in the sequence alignment. GenBank Accession numbers are in Fig. 3.

rapid vertical inactivation, without a significant increase in their copy numbers during the evolution after the initial transpositions and multiplications in the genome. A similar conclusion was drawn in a recent study in the *mellifera* subfamily of MLE (Lampe *et al*., 2003), suggesting that the selection upon the *mariner* was primarily at the horizontal transfer between genomes.

One of the major mechanisms of vertical inactivation is large deletions (> 10 bp) containing short direct repeat (5–7 bp) sequences near the end of the deletions (Fig. S3). Eighty percent of the large deletions were identified for the presence of short direct repeats at or near the end of the deletions. *Drosophila Mos-1* elements were also found to have a similar phenomenon, having vertical inactivations by large deletions with short direct repeats at the end of the deletions (Brunet *et al*., 2002). In our MLE survey, only three clones for each 5′ and 3′ ITR were clearly determined for their ITR sequences (Fig. 2), whereas for other sequences, four and three clones for each 5′ and 3′ ITR seemed to be lost by deletions. Thus, more than 50% of MLE in PBW may have lost their ITR sequence as a process of vertical inactivation that prevents mobilization.

MLE in different populations of PBW

Patterns of polymorphism in MLE among different populations may provide an evolutionary history of the MLE and also of its host PBW. Because PBW was introduced to a new continent by seed shipments, we focused on an examination of the differences between Indian and US populations. It was surprising that only a few variations were observed among different strains in genomic Southern blotting and TE displays for the banding patterns (Fig. 4). Copy number of the MLE can be roughly estimated as 6–10 copies per diploid genome, based on the number of bands on the Southern blotting. In the genomic Southern, the strong bands 1.8, 2, 4 and 7 (Fig. 4, band ID) are fixed in all individuals we tested, whereas the weak bands between 2.6 and 3.7 are polymorphic (Fig. 4). Sequences for the core fragment of the MLE from the Akola collection from India (MLEF1b and MLEF1d in Fig. 1) share identical sequences with

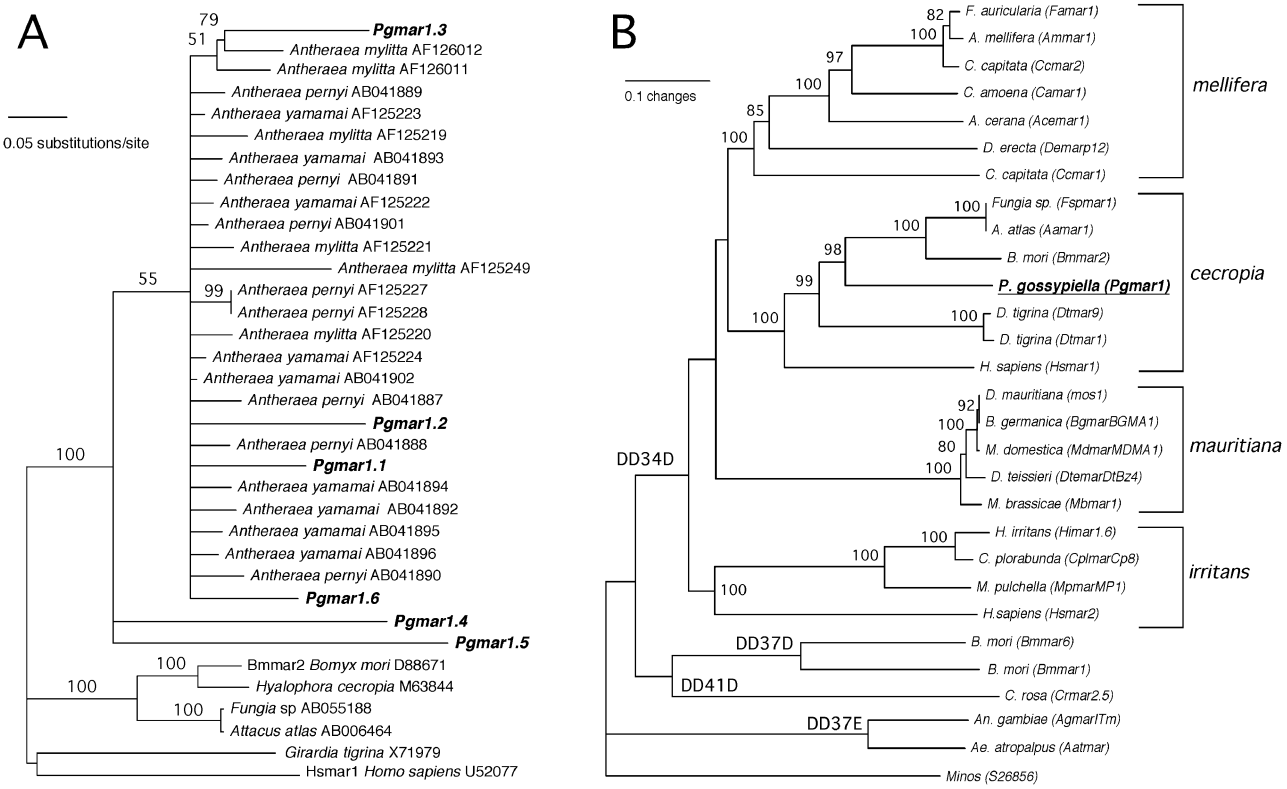

Figure 3. Phylogenetic relationship of *P. gossypiella mariner*-like nucleotide sequences (*Pgmar1.1–1.6* in bold letters) with other *cecropia mariners* (A). Phylogram for amino acid sequences for the putative transposases of *mariner*-like elements (MLE) (B). Trees were generated by the distance method. Numbers on the nodes are the bootstrapping values in 1000 replications. The MLE subfamilies are grouped on the right column. GenBank Accession numbers are: *Famar1* (AAO12863), *Ammar1* (AAO12861), *Ccmar2* (AO12864), *Camar1* (AAO12862), *Acemar1* (BAB86288), *Demarp12* (U08093), *Ccmar1* (AAB17945), *Fspmar1* (BAB32436), *Aamar1* (BAA21826), *Bmmar2* (BAA23532), *Dtmar9* (CAA56856), *Dtmar1* (CAA50801), *Hsmar1* (AAC52010), *Mos 1* (A26491), *BgmarBGMA1* (AAK40118), *MdmarMDMA1* (AAK54758), *DtemarDtBz4* (AAC28262), *Mbmar1* (AAL69970), *Himar1.6* (U11645), *CplmarCp8* (AAC46948), *MpmarMP1* (U11649), *Hsmar2* (AAC52011), *Bmmar6* (AAN06610), *Bmmar1* (U47917), *Crmar2.5* (AAK61417), *AgmarITm* (AAL16724), *Aatmar* (AAL16723), and *Minos* (S26856). The *Pgmar* sequences are deposited in GenBank with Accession numbers: *Pgmar1.1* (DQ023493), *Pgmar1.2* (DQ023494), *Pgmar1.3* (DQ023495), *Pgmar1.4* (DQ023496), *Pgmar1.5* (DQ023497), *Pgmar1.6* (DQ023498), *5MLEd* (DQ023491), *5MLEg* (DQ023492), *3MLEa* (DQ023488), *3MLEb* (DQ023489), and *3MLEe* (DQ023490).

those from other US strains, APHIS and ARS. Although our survey was not exhaustive, these observations together suggest that the MLE in PBW are ancient and are in the process of accumulation of inactivating mutations that may consequently result in stochastic loss of the MLE.

Experimental procedures

Insect strains

Adult pink bollworms APHIS and ARS strains were maintained in the USDA-APHIS pink bollworm rearing facility (Phoenix, AZ). Two laboratory strains, SAF-H and MOV-H, and two field populations from India collected from Akola and Kallakal, were kindly provided by Bruce Tabashnik (University of Arizona, Tucson, AZ).

General approaches

To obtain MLE sequences, we amplified a partial fragment of MLE by using a pair of degenerate primers. The sequence information from the fragments was used for inverse PCR that exposed an ITR. Then the PCR with the primers designed on the ITR (ITR-PCR) pulled out several different copies of MLE. Diversity of the MLE among different populations was surveyed by using genomic Southern hybridizations and TE display. Flanking sequences revealed in the TE display were used for designing primers on the flanking sequence. PCR amplification of whole contiguous MLE by using combinatorial pairs of primers on each flanking sequence was performed.

Genomic DNA isolation PCR, cloning and sequencing MLEs

Genomic DNA (gDNA) was extracted from single adult or larval PBW. Individual PBW were homogenized in 400 μl of lysis buffer (10 mM Tris-HCl pH 8.0, 10 mM EDTA pH 8.0, 0.5% SDS and 200 mM NaCl). The homogenate was treated with 50 μg/ml of RNase A (Sigma, St Louis, MO) for 30 min at 37 °C, then incubated with proteinase K (100 μg/ml) at 55 °C for 3 h. Three phenol-chloroform extractions were followed by ethanol precipitation of the DNA. PCR was performed in a 25 μl reaction volume containing 20–50 ng gDNA, 0.8 μM of each degenerate primer, 0.2 mM of

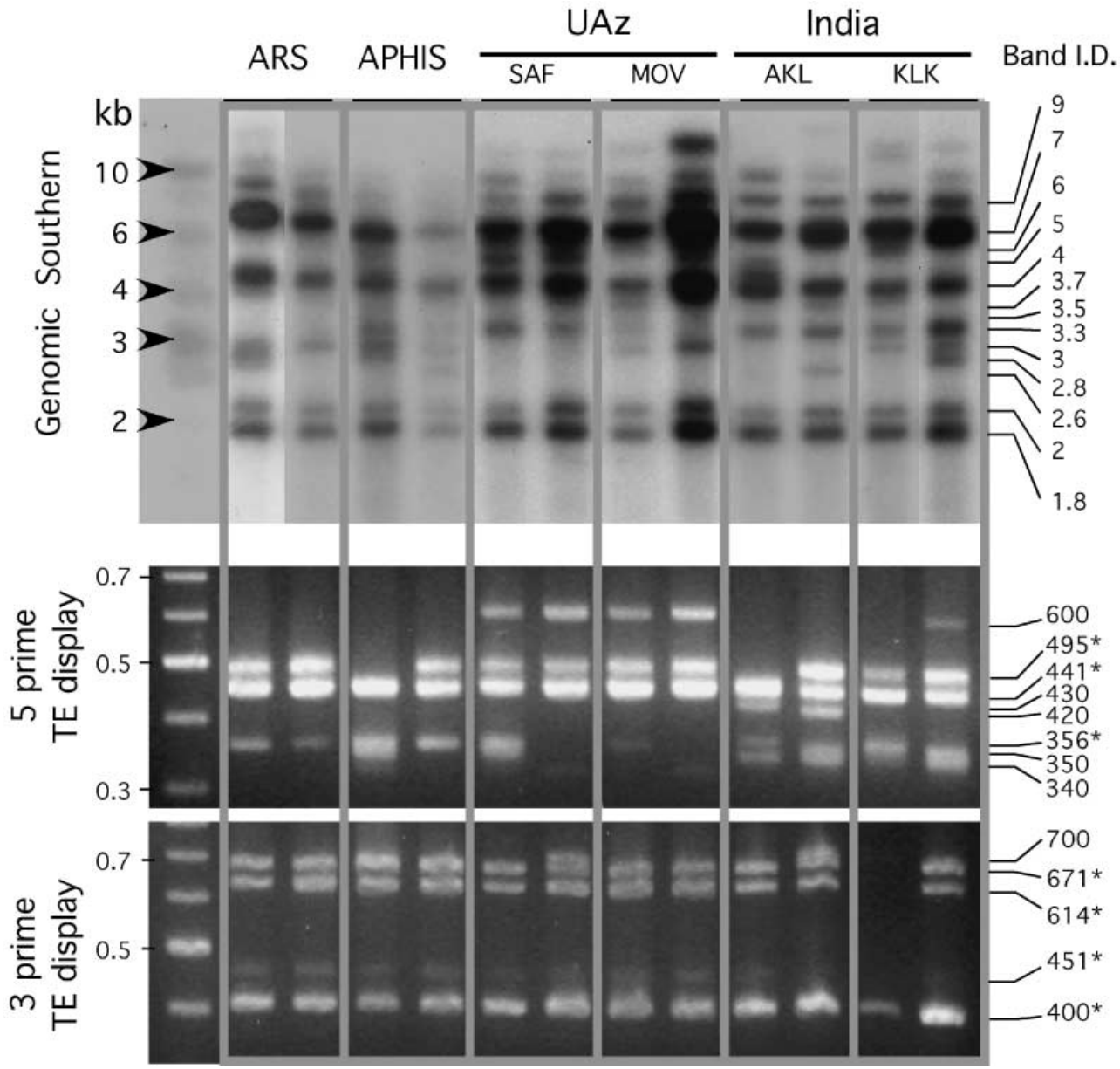

Figure 4. Southern hybridization of *Eco*RI digested individual PBW gDNA from six different strains (upper panel). The 5′ and 3′ TE displays (middle and lower panel, respectively) by using vectorette PCR for individual PBW. Same set of individuals shown in Southern bot was used for the TE displays. Southern blot result is presented by combining the results of multiple films with different exposure duration for the optimal band strengths for each lane. Bands were named (Band ID) for their approximate size. The Band IDs with asterisk (*) are the size matching the sequenced clones obtained from the APHIS strain.

each dNTP, 2 mM of $MgCl_2$, and 0.5 unit of REDTaq Genomic DNA Polymerase (Sigma) in the buffer supplied by manufacturer (100 mM Tris-HCl, pH 8.3, 500 mM KCl, and 0.1% gelatin). MLE in PBW were amplified by using a pair of degenerate primers described in Robertson (1993): MAR124F (5′-TGGGTNCCNCAYGARYT-3′) and MAR276R (5′-GGNGCNARRTCNGGNSWRTA-3′) (Fig. 1). Touchdown PCR for the degenerate PCR consisted of denaturing gDNA at 94 °C for 5 min, then 20 cycles at 94 °C for 30 s, 50–40.5 °C (decreasing by −0.5 °C per cycle) for 30 s, 72 °C for 1 min; 15 cycles at 94 °C for 30 s, 40 °C for 30 s, 72 °C for 1 min; and a final extension at 72 °C for 10 min. Based on sequence information obtained from the amplified product of degenerate PCR, two pairs of primers were designed for nested inverse PCR. Genomic DNA (~2 μg) was digested with each 6-cutter enzyme, *Eco*RI, *Xho*I, *Hin*dIII or *Pst*I (Promega, Madison, WI), for 10–16 h at 37 °C in each 20 μl reaction. The digestions were followed by self-ligation using T4 DNA-ligase and precipitation, and were used as the template for PCR. The first-round PCR was carried out with external primers (MEF: 5′-TGGAGAAGCTGGCACATCCC-3′; MER: 5′-CGGTATGCGGTTCCAAATAC-3′), then second-round PCR was

carried out by using internal primers (MIF: 5′-TTAATCGCTCGTCTC-CACTGC-3′ and MIR: 5′-AGGCACGTCTTGCACAGAAT-3′). The full-length MLE were obtained by PCR amplification with the primer designed on the 3′ ITR (5′-GTCGGTCAGCGCAACGAC-GACGCT-3′). Flanking-sequence PCR for isolating contiguous MLE were performed with primers designed on each flanking sequences captured in the vectorette PCR. The pairs provided successful amplifications of *mariner* were MCF1: 5′-TTCTTCGCT-TCATTCATCAG-3′ and MCR1 : 5′-CCAATGACCACGAAAGAAGT-3′ for *Pgmar1.1* and MCF2: 5′-CTACCCAGATTATTGCTTCG-3′ and MCR2: 5′-TTTTATTTCCAGTAAGTAGGTAC-3′ for *Pgmar1.6*. Sequence alignment was performed with CLUSTAL X (Thompson *et al.*, 1994). The aligned sequences were used for construction of phylogenetic trees in PAUP 4b.2. (Swofford, 2000).

Vectorette PCR for TE display

We employed vectoretted PCR to obtain the flanking sequence of MLE (Ko *et al.*, 2003). The vectorette sequence and vectorette construction were modified from Ko *et al.* (2003). Two anchoring bubble linker oligoes were designed to make the vectorette unit for ligation to the *Bfa*I digested gDNA:

```
vect 53 5′  CCCTTCTCGAATCGTAACCGTTCGGTCCTCTG 3′
            ||||||||    ||     |  | | ||||||||
vect 57 3′  GGGAAGAGAGCAGGCAAGAAATGGCAGGAGACAT 5′
```

Digestions were performed at 37 °C overnight by using *Bfa*I (NEB, Beverly, CA) in a 25 µl reaction containing 2.5 µl 10× OPA buffer (One-Phor-All Buffer *PLUS*, Amersham, Piscataway, NJ), 1 µg DNA, and 10 units of *Bfa*I. After digestion, 3 pmol of anchor bubble unit, 50 nmol ATP (Sigma), and 6 units of T4 DNA ligase (Promega) were added, and the reaction was incubated at 16 °C for 16 h.

Two consecutive rounds of nested PCR with two sets of primers were carried out. The primers used for the vectorette PCR are:

TEMR1 : 5′-TGTAGTGTTAGCTCCGTAAACG-3′ (5′ specific external MLE primer)

TEMF1 : 5′-CACAGCAGACGGTCTCCAAG-3′ (3′ specific external MLE primer)

TEMR2 : 5′-GCTGCCACGACGGTACTCAT-3′ (5′ specific internal MLE primer)

TEMF2 : 5′-CTTGCGCCGACAGACTACTA-3′ (3′ specific internal MLE primer)

VPCR1 : 5′-CCCTTCTCGAATCGTAACCG-3′ (vectorette external primer)

VPCR2 : 5′-AACCGTTCGGTCCTCTGTAG-3′ (vectorette internal primer)

The PCR products were separated on 2.4% agarose gel and visualized by ethidium bromide staining.

Southern blot hybridization

Individual gDNA were digested with *Eco*RI, run on 0.8% agarose gel, transferred to a nylon membrane (Hybond-N^+, Amersham), and cross-linked by UV light in UVP HL-2000 Hybrilinker. The 463 bp MLEF1d fragment (Fig. 1) was labelled with [^{32}P]dCTP as a probe by random priming with the Megaprime™ DNA labelling system (Amersham). Hybridizations were performed at 65 °C in Rapid-Hyb Buffer (Amersham). After the hybridization, membranes were washed twice, 10 min each, in 2× SSC, 0.1% SDS wash solution at room temperature, then washed twice at 65 °C in 0.5× SSC and 0.1% SDS wash solution for 15 min each.

Acknowledgements

We thank Bruce Tabashnik for providing four pink bollworm strains, Gregory Simmons and Ernie Miller for providing APHIS and ARS strains, Xiaoxia Ren for valuable comments, and Kun Yan Zhu for reviewing an earlier version of this manuscript. This paper is contribution no. 05-274-J from Kansas Agricultural Experiment Station, Manhattan, Kansas. This work was supported by USDA-APHIS Cooperative agreement award no. 03-8100-0879-CA.

References

Brunet, F., Giraud, T., Godin, F. and Capy, P. (2002) Do deletions of *mos1*-like elements occur randomly in the drosophilidae family? *J Mol Evol* **54**: 227–234.

Fryxell, K.J. and Miller, T.A. (1995) Autocidal biological control: a general strategy for insect control based on genetic transformation with a highly conserved gene. *J Econ Entomol* **88**: 1221–1232.

Holt, R.A., Subramanian, G.M., Halpern, A., Sutton, G.G., Charlab, R., Nusskern, D.R., *et al.* (2002) The genome sequence of the malaria mosquito *Anopheles gambiae*. *Science* **298**: 129–149.

Jacobson, J.W., Medhora, M.M. and Hartl, D.L. (1986) Molecular structure of a somatically unstable transposable element in *Drosophila*. *Proc Natl Acad Sci USA* **83**: 8684–8688.

Jarvik, T. and Lark, K.G. (1998) Characterization of *Soymar1*, a *mariner* element in soybean. *Genetics* **149**: 1569–1574.

Kapitonov, V.V. and Jurka, J. (2003) Molecular paleontology of transposable elements in the *Drosophila melanogaster* genome. *Proc Natl Acad Sci USA* **100**: 6569–6574.

Ko, W.Y., David, R.M. and Akashi, H. (2003) Molecular phylogeny of the *Drosophila melanogaster* species subgroup. *J Mol Evol* **57**: 562–573.

Lampe, D.J., Witherspoon, D.J., Soto-Adames, F.N. and Robertson, H.M. (2003) Recent horizontal transfer of *mellifera* subfamily *mariner* transposons into insect lineages representing four different orders shows that selection acts only during horizontal transfer. *Mol Biol Evol* **20**: 554–562.

Lander, E.S., Linton, L.M., Birren, B., Nusbaum, C., Zody, M.C., Baldwin, J. *et al.* (2001) Initial sequencing and analysis of the human genome. *Nature* **409**: 860–921.

Langin, T., Capy, P. and Daboussi, M.J. (1995) The transposable element *impala*, a fungal member of the *Tc1-mariner* superfamily. *Mol Gen Genet* **246**: 19–28.

Lidholm, D.A., Gudmundsson, G.H. and Boman, H.G. (1991) A highly repetitive, *mariner*-like element in the genome of *Hyalophora cecropia*. *J Biol Chem* **266**: 11518–11521.

Lohe, A.R., Moriyama, E.N., Lidholm, D.A. and Hartl, D.L. (1995) Horizontal transmission, vertical inactivation, and stochastic loss of *mariner*-like transposable elements. *Mol Biol Evol* **12**: 62–72.

Mandrioli, M. (2000) *Mariner*-like transposable elements are interspersed within the rDNA-associated heterochromatin of the pufferfish *Tetraodon fluviatilis* (Osteichthyes). *Chromosome Res* **8**: 177–179.

Nakajima, Y., Fujimoto, H., Negishi, T., Hashido, K., Shiino, T., Tsuchida, K., *et al.* (2002) Possible horizontal transfer of *mariner*-like sequences into some invertebrates including lepidopteran insects, a grasshopper and a coral. *J Insect Biotechnol Sericol* **71**: 109–121.

Peloquin, J.J., Thibault, S.T., Staten, R. and Miller, T.A. (2000) Germ-line transformation of pink bollworm (Lepidoptera: gelechiidae) mediated by the *piggyBac* transposable element. *Insect Mol Biol* **9**: 323–333.

Robertson, H.M. (1993) The *mariner* transposable element is widespread in insects. *Nature* **362**: 241–245.

Robertson, H.M. and Zumpano, K.L. (1997) Molecular evolution of an ancient *mariner* transposon, *Hsmar1*, in the human genome. *Gene* **205**: 203–217.

Swofford, D.L. (2000) *PAUP*. Phylogenetic Analysis Using Parsimony (*and Other Methods), Version 4.* Sinauer Associates, Sunderland, Massachusetts.

Thomas, D.D., Donnelly, C.A., Wood, R.J. and Alphey, L.S. (2000) Insect population control using a dominant, repressible, lethal genetic system. *Science* **287**: 2474–2476.

Thompson, J.D., Higgins, D.G. and Gibson, T.J. (1994) CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids Res* **22**: 4673–4680.

Supplementary material

The following supplementary material is available for this paper online:

Figure S1. Sequence alignment of multiple copies of *mariner*-like elements obtained from pink bollworm. The region encoding transposase is aligned with consensus sequence for 50% majority rule. Ancestral sequence was reconstructed by following the consensus sequence, ignoring deleted or inserted sequences in the calculation of consensus (bold). Bottom line is the conceptual translation of the ancestral sequence.

Figure S2. Deletions and insertions found in the sequence alignment among *Pgmars*. Locations of the indels are indicated by nucleotide numbering. Deletions are indicated by open triangles and insertions by solid inverted triangles. The bottom is a histogram showing distribution of the sizes of insertions and deletions.

Figure S3. Large deletions containing short direct repeats at or near the end of deletions. Underlined, capitalized and italicized letters are presumed perfect or imperfect short direct repeats (5–7 bp). The sequence on the upper lines is for the consensus sequence and the bottom line is for the sequence containing the deletion.